PART 1

Hamiltonian dynamics and symplectic geometry

1

The least action principle and Hamiltonian mechanics

In this chapter, we start by reviewing Hamilton's least action principle in classical mechanics. We will motivate all the basic concepts in symplectic geometry out of this variational principle. We refer the reader to (Go80) or (Ar89) for further physical applications of this principle.

1.1 The Lagrangian action functional and its first variation

We start with our explanations on $\mathbb{R}^n$ and then move onto general configuration space M. We call $\mathbb{R}^n$ the *(configuration) space* and $\mathbb{R} \times \mathbb{R}^n$ the *space-time*. We denote by q^i, $i = 1, \ldots, n$ the standard coordinate functions of $\mathbb{R}^n$, and by

$$(q^1, \ldots, q^n, v^1, \ldots, v^n)$$

the associated *canonical coordinates* of $T\mathbb{R}^n \cong \mathbb{R}^n \times \mathbb{R}^n$. It is customary to denote the canonical coordinates by

$$(q^1, \ldots, q^n, \dot{q}^1, \ldots, \dot{q}^n)$$

instead, especially in the physics literature. We will follow this convention, whenever there is no danger of confusion.

We denote by $\gamma : [t_0, t_1] \to \mathbb{R}^n$ a continuous path, regarded as the trajectory of a moving particle. In coordinates, we may write

$$\gamma(t) = (q^1(t), \ldots, q^n(t)), \quad q^i(t) = q^i(\gamma(t)).$$

We say $[t_0, t_1]$ is the domain of the path and denote

$$\mathcal{P} = \mathcal{P}_{t_0}^{t_1}(\mathbb{R}^n) = \{\gamma \mid \mathrm{Dom}(\gamma) = [t_0, t_1]\}.$$

In the Lagrangian formalism of classical mechanics, the relevant action functional, called the *Lagrangian action functional*, has the form

$$\Phi(\gamma) = \int_{t_0}^{t_1} L(t, \gamma, \dot\gamma) dt$$

where L is a function, called the *Lagrangian*,

$$L = L(t, q, \dot q) : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}.$$

Example 1.1.1 Consider a motion on $\mathbb{R}^n$ over the time interval $[t_0, t_1]$, i.e., a map $\gamma : [t_0, t_1] \to \mathbb{R}^n$.

(i) The *energy functional* is defined by

$$E(\gamma) = \frac{1}{2} \int_{t_0}^{t_1} |\dot\gamma|^2 \, dt.$$

(ii) The length of the path $\gamma : [t_0, t_1] \to \mathbb{R}^n$ is given by

$$L(\gamma) = \int_{t_0}^{t_1} |\dot\gamma| dt.$$

As in the mechanics literature, we denote by $\Delta\gamma$ the *infinitesimal variation*. In the more formal presentation, we note that the tangent space of $\mathbb{R}^n$ at a given point x is canonically identified with $\mathbb{R}^n$ itself. Therefore we can write a variation $h = \Delta\gamma \in T_\gamma\mathcal{P}$, the tangent space, at the path γ as a map

$$h : [t_0, t_1] \to \mathbb{R}^n.$$

We denote by $|h|$ the norm of h with respect to a given norm on the linear space

$$T_\gamma \mathcal{P}_{t_0}^{t_1}(\mathbb{R}^n) \cong \Gamma(\gamma^*(T\mathbb{R}^n)).$$

We will not delve into the matter of giving the precise mathematical description of the following definition, which is the analog of the standard definitions to the finite dimensional case in the present *infinite-dimensional* case. We refer to e.g., (AMR88) for the precise mathematical definitions.

Definition 1.1.2 Let Φ be as above.

(1) A functional Φ is *differentiable* at x if there exists a linear map $F(\gamma)$ such that

$$\Phi(\gamma + \Delta\gamma) - \Phi(\gamma) = F(\gamma) \cdot \Delta\gamma + R, \tag{1.1.1}$$

where $R = R(\gamma, h) = O(|h|^2)$.

(2) If this holds, $F(\gamma)$ is said to be the *differential* of Φ at γ and denoted by $d\Phi(\gamma)$.
(3) We call any path γ an *extremal path* if it satisfies $F(\gamma) \cdot h = 0$ for all variation h.

Now we find the formula for the differential $F(\gamma) \cdot h$ in terms of the Lagrangian density L. For given $h : [t_0, t_1] \to \mathbb{R}^n$ regarded as a variation (or a tangent vector) of γ, we have

$$\begin{aligned} F(\gamma) \cdot h &= \left.\frac{d}{ds}\right|_{s=0} \Phi(\gamma + sh) \\ &= \left.\frac{d}{ds}\right|_{s=0} \int_{t_0}^{t_1} L(t, q + sh, \dot{q} + s\dot{h}) dt \\ &= \int_{t_0}^{t_1} \left.\frac{d}{ds}\right|_{s=0} L(t, q + sh, \dot{q} + s\dot{h}) dt \\ &= \int_{t_0}^{t_1} \left(\frac{\partial L}{\partial q} \cdot h + \frac{\partial L}{\partial \dot{q}} \cdot \dot{h} \right) dt. \end{aligned}$$

We integrate the second term by parts and get

$$\int_{t_0}^{t_1} \frac{\partial L}{\partial \dot{q}} \cdot \dot{h}\, dt = \frac{\partial L}{\partial \dot{q}} \cdot h \Big|_{t_0}^{t_1} - \int_{t_0}^{t_1} \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) \cdot h\, dt.$$

Therefore,

$$F(\gamma) \cdot h = \int_{t_0}^{t_1} \left(\frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right)\right) \cdot h\, dt + \frac{\partial L}{\partial \dot{q}} \cdot h \Big|_{t_0}^{t_1}. \tag{1.1.2}$$

To describe the condition of extremal paths of Φ in terms of the Lagrangian density function L, we require that the boundary term, i.e., the second term of (1.1.2), vanish. There are two common ways of achieving this goal in the mechanics.

1. Two-point boundary condition. We define the subset

$$\mathcal{P}_{t_0}^{t_1}(\mathbb{R}^n; q_0, q_1) = \{\gamma \in \mathcal{P}_{t_0}^{t_1}(\mathbb{R}^n) \mid q(t_0) = q_0, \quad q(t_1) = q_1\} \tag{1.1.3}$$

of $\mathcal{P}_{t_0}^{t_1}(\mathbb{R}^n)$, which consists of the paths satisfying the so-called *two-point boundary condition*. In this case, the variation $h = \Delta\gamma$ satisfies

$$h(t_0) = 0 = h(t_1).$$

For such γ and h, we have

$$\frac{\partial L}{\partial \dot{q}} \cdot h \Big|_{t_0}^{t_1} = \frac{\partial L}{\partial \dot{q}}(t_1, \gamma(t_1), \dot{\gamma}(t_1)) \cdot h(t_1) - \frac{\partial L}{\partial \dot{q}}(t_0, \gamma(t_0), \dot{\gamma}(t_0)) \cdot h(t_0) = 0 - 0 = 0.$$

Therefore, if we restrict Φ to this subset $\mathcal{P}_{t_0}^{t_1}(\mathbb{R}^n; q_0, q_1)$, the corresponding restricted functional has the first variation given by

$$F(\gamma) \cdot h = \int_{t_0}^{t_1} \left(\frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right)\right) \cdot h\, dt. \tag{1.1.4}$$

Hence we have derived the *equation of motion*.

Proposition 1.1.3 *Let q_0 and q_1 be two fixed points in $\mathbb{R}^n$. Consider the functional*

$$\Phi(\gamma) = \int_{t_0}^{t_1} L(t, \gamma, \dot{\gamma}) dt$$

for the paths γ under the two-point boundary condition

$$\gamma(0) = q_0, \quad \gamma(1) = q_1.$$

Then γ is extremal if and only if it satisfies

$$\frac{\partial L}{\partial q^i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) = 0, \quad i = 1, \ldots, n. \tag{1.1.5}$$

We next discuss another natural boundary condition, the periodic boundary condition.

2. Periodic boundary conditions. Consider the subset

$$\mathcal{L}_{t_0}^{t_1}(\mathbb{R}^n) = \{\gamma \in \mathcal{P}_{t_0}^{t_1}(\mathbb{R}^n) \mid \gamma(t_0) = \gamma(t_1),\ \dot{\gamma}(t_0) = \dot{\gamma}(t_1)\}.$$

The corresponding variation $h = \Delta$ will satisfy the *periodic boundary condition*

$$h(t_0) = h(t_1).$$

In addition, *provided the density function $L = L(t, q, \dot{q})$ satisfies the time-periodic condition*

$$L(t_0, \cdot, \cdot) \equiv L(t_1, \cdot, \cdot), \tag{1.1.6}$$

the boundary term of (1.1.2) again vanishes; this time, however, because we have

$$\frac{\partial L}{\partial \dot{q}}(t_1, \gamma(t_1), \dot{\gamma}(t_1)) \cdot h(t_1) = \frac{\partial L}{\partial \dot{q}}(t_0, \gamma(t_0), \dot{\gamma}(t_0)) \cdot h(t_0).$$

We summarize this as follows.

Proposition 1.1.4 *Suppose L satisfies (1.1.6). Consider the functional*

$$\Phi(\gamma) = \int_{t_0}^{t_1} L(t, \gamma, \dot{\gamma}) dt$$

restricted to the paths γ under the periodic boundary condition

$$\gamma(t_0) = \gamma(t_1), \quad \dot\gamma(t_0) = \dot\gamma(t_1).$$

Then γ is extremal if and only if it satisfies (1.1.5).

Equation (1.1.5) is called the *Euler–Lagrange equation* of L. Note that this is the *second-order ODE* with respect to the variable $\gamma(t) = (q^1(t), \ldots, q^n(t))$.

Remark 1.1.5 Since the above discussion is independent of the choice of coordinates $(q^1, \ldots, q^n)$, *as long as we use the associated canonical coordinates* of $T\mathbb{R}^n \cong \mathbb{R}^n \times \mathbb{R}^n$, the Euler–Lagrange equation for L is also coordinate-independent (or *covariant* in the physics terminology). In other words, if $(Q^1, \ldots, Q^n)$ is another coordinate system of $\mathbb{R}^n$, then the associated Euler–Lagrange equation has the same form as

$$\frac{\partial L}{\partial Q_i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot Q_i}\right) = 0, \quad i = 1, \ldots, n.$$

1.2 Hamilton's action principle

The Lagrangian that is relevant in Newtonian mechanics has the form

$$L = T - U, \tag{1.2.7}$$

where T is the *kinetic energy*

$$T = T(x, \dot x) = \frac{1}{2} m \dot x \cdot \dot x$$

and $U = U(x) : \mathbb{R}^n \to \mathbb{R}$ is the *potential energy*, which is a function depending only on the position vector $x \in \mathbb{R}^n$. Then Newton's equation of motion

$$\frac{d}{dt}(m\dot q^i) = -\frac{\partial U}{\partial q^i}, \quad i = 1, \ldots, n, \tag{1.2.8}$$

is equivalent to the Euler–Lagrange equation (1.1.5) for the Lagrangian (1.2.7). This gives rise to the following:

Hamilton's least action principle. Motions of the mechanical system under Newton's second law coincide with the extremals of the functional

$$\Phi(\gamma) = \int_{t_0}^{t_1} L\, dt$$

(*under an appropriate boundary condition as mentioned before*).

Recall that, in Newtonian mechanics, the *momentum vector* is defined by

$$m\dot{x} =: p \tag{1.2.9}$$

and the *force field* is provided by

$$-\nabla U =: F.$$

Owing to the special form of the Lagrangian given in (1.2.7), $p_i = m\dot{q}^i$ is nothing but $\partial L/\partial \dot{q}^i$. This leads to the following notion in classical mechanics. (See (Go80) or (Ar89) for more discussion.)

Definition 1.2.1 Let L be an arbitrary Lagrangian. Then the *generalized momenta*, denoted as p_i, are defined by

$$p_i = \frac{\partial L}{\partial \dot{q}^i}. \tag{1.2.10}$$

With this definition, the Euler–Lagrange equation becomes

$$\dot{p}_i = \frac{\partial L}{\partial q^i}, \quad i = 1, \ldots, n. \tag{1.2.11}$$

Exercise 1.2.2 Interpret (1.2.10) and (1.2.11) in the invariant fashion. Explain why one should regard the right-hand side thereof as a covariant one-tensor or as a differential one-form.

1.3 The Legendre transform

In solving a mechanical problem, one often first finds the formula for the momenta p_i in time and then would like to convert this into a formula for the position coordinates q^i. This is not always possible, though. A necessary condition for the Lagrangian L is its *convexity* with respect to $\dot{x}$ for each fixed position vector x. Such a function should be considered as a family of convex functions

$$\dot{x} \mapsto L(t, x, \dot{x}); \mathbb{R}^n \to \mathbb{R}$$

parameterized by the position vector $(t, x) \in \mathbb{R} \times \mathbb{R}^n$.

In this section, we discuss an important operation, called the *Legendre transform*, that appears in many branches of mathematics. The Legendre transform recently played an important role in a rigorous formulation of the mirror symmetry in relation to the Strominger–Yau–Zaslow proposal. We refer the reader to (SYZ01), (Hi99) and (GrSi03) for more explanation of this aspect. Partly because of this recent resurgence of interest, we provide some detailed mathematical explanations of the Legendre transform in an invariant fashion. After that, we will return to the Hamiltonian formulation of the classical mechanics.

1.3.1 The Legendre transform of a function

Let V be a (finite-dimensional) vector space and V^* its dual vector space. We denote by $\langle\ ,\ \rangle$ the canonical paring between V and V^*.

Definition 1.3.1 Let $U \subset V$ be an open subset. A function $f : V \to \mathbb{R}$ is said to be *convex* on $U \subset V$ if it satisfies

$$f((1-t)x_1 + tx_2) \le (1-t)f(x_1) + tf(x_2) \tag{1.3.12}$$

for all $t \in [0, 1]$ and for all $x_1, x_2 \in U$, and *strictly convex* if

$$f((1-t)x_1 + tx_2) < (1-t)f(x_1) + tf(x_2) \tag{1.3.13}$$

for all $t \in (0, 1)$ and for all $x_1, x_2 \in U$.

The following is an easy exercise to prove.

Lemma 1.3.2
(1) *Any convex function f on U is continuous on U.*
(2) *Any strictly convex function $f : V \to \mathbb{R}$ that is bounded below has the unique minimum point if it has one.*

Example 1.3.3 Let $V = \mathbb{R}$ and consider the function $f(x) = x^\alpha/\alpha$ with $\alpha > 1$. Then f is convex on $\mathbb{R}$.

For a given function $f : V \to \mathbb{R}$, we consider the function $F : V \times V^* \to \mathbb{R}$ defined by

$$F(x, p) := \langle x, p\rangle - f(x)$$

and the value

$$g(p) = \sup_{x\in V} F(x, p). \tag{1.3.14}$$

The new function g, if defined, is called the *Legendre transform* or the *Fenchel transform*.

In general, the value of g need not be finite. However, whenever the value is defined, we have the inequality

$$\langle x, p\rangle \le f(x) + g(p), \tag{1.3.15}$$

Cambridge University Press
978-1-107-07245-9 - Symplectic Topology and Floer Homology: Volume 1: Symplectic Geometry and Pseudoholomorphic Curves
Yong-Geun Oh
Excerpt
More information

which is called the *Fenchel inequality*. To make the value of g finite everywhere, one needs to impose the following superlinearity of f.

Definition 1.3.4 Let V be a (finite-dimensional) vector space. A function $f : V \to \mathbb{R}$ is said to be superlinear, if f is bounded below and

$$\lim_{|x|\to\infty} \frac{f(x)}{|x|} = +\infty. \tag{1.3.16}$$

Exercise 1.3.5 Prove that the superlinearity in this definition is equivalent to the statement that, for all $K < \infty$, there exists $C(K) > -\infty$ such that $f(x) \geq K|x| + C(K)$ for all $x \in V$.

An example of a convex but not superlinear function is $f(x) = e^{-x}$ as a function on $V = \mathbb{R}$.

We borrow the following from Proposition 1.3.5 of Fathi's book (Fa05) restricted to the finite-dimensional cases.

Proposition 1.3.6 *Let $f : V \to \mathbb{R}$ be a function. Then the following apply.*

(1) *If f is superlinear, then g is finite everywhere.*
(2) *If g is finite everywhere, it is convex.*
(3) *If f is continuous, then g is superlinear.*

Proof We first prove (1). By the superlinearity (1.3.16) and Exercise 1.3.5, there exists a constant $C = C(|p|)$ such that $f(x) \geq |p||x| + C(|p|)$ for all $x \in V$. Therefore we have

$$\langle x, p\rangle - f(x) \leq |x||p| - f(x) \leq |x||p| - (|x||p| + C(|p|)) = -C(|p|) < \infty.$$

This proves $g(p) = \sup_{x\in V}(\langle x, p\rangle - f(x)) < \infty$. On the other hand, we have

$$g(p) = \sup_{x\in V}(\langle x, p\rangle - f(x)) \geq -f(0) > -\infty,$$

which proves (1). For the property (2), we note that g is the upper bound for the family of linear functions, which is obviously convex,

$$p \mapsto \langle x, p\rangle - f(x),$$

and hence g must be convex. To show (3), we will apply Exercise 1.3.5. For any K, we derive

$$g(p) \geq \sup_{|x|=K} \langle x, p\rangle - \sup_{|x|=K} f(x)$$